Vector-Neuron Models of Associative Memory

Boris V. Kryzhanovsky
Institute
of Optical Neural Technologies
Russian Academy of Sciences
Moscow, Russia
E-mail: kryzhanov@iont.ru

Leonid B. Litinskii
Institute
of Optical Neural Technologies
Russian Academy of Sciences
Moscow, Russia
E-mail: litin@iont.ru

Andrey L. Mikaelian
Institute
of Optical Neural Technologies
Russian Academy of Sciences
Moscow, Russia
E-mail: iont@postman.ru

***Abstract*— We consider two models of Hopfield-like associative memory with q-valued neurons: Potts-glass neural network (PGNN) and parametrical neural network (PNN). In these models neurons can be in more than two different states. The models have the record characteristics of its storage capacity and noise immunity, and significantly exceed the Hopfield model. We present a uniform formalism allowing us to describe both PNN and PGNN. This networks inherent mechanisms, responsible for outstanding recognizing properties, are clarified.**

I. Introduction

The number of patterns that can be stored in Hopfield model (HM) is comparatively not so large. If N is binary neurons number, then the thermodynamic approach leads to the well-known estimation of the HM storage capacity, $p_{HM} \sim 0.14 \cdot N$ ([1], [2]). At the early 90-th some authors suggested Hopfield-like models of associative memory with q-valued neurons that can be in more than two different states, $q \geq 2$, [3]-[10]. All these models are related with the Potts model of magnetic. The last one generalizes the Ising model for the case of the spin variable that takes $q > 2$ different values [11],[12]. In all these works the authors used the same well-known approach linking the Ising model with the Hopfield model (see, for example, [2]). Namely, in place of the short-range interaction between two nearest spins the Hebb type interconnections between all q-valued neurons were used. As a result, long-range interactions appear. Then in the mean-field approximation it was possible to calculate the statistical sum and, consequently, to construct the phase diagram. Different regions of the phase diagram were interpreted in the terms of the network ability to recognize noisy patterns.

For all these models, except one, the storage capacity is even less than that for HM. An exception is so named *Potts-glass neural network* (PGNN) [3]. The numerical solution of transcendential equation system resulting from thermodynamic approach leads to the following estimation for storage capacity for PGNN

$$p_{PGNN} \sim \frac{q(q-1)}{2} \cdot p_{HM}.$$

As far as q-valued models are intended for color images processing, number q stands for number of different colors, used for elementary pixel can be painted. Even if $q \sim 10$ the storage capacity of PGNN is 50 times as much as the storage capacity of HM. For computer processing of colored images the standard value is $q = 256$. Consequently, comparing with HM the gain is about four orders, $p_{PGNN} \sim 10^4 \cdot p_{HM}$. It is very good result. However, for long time it was not clear, why PGNN has such a big storage capacity. Thermodynamic approach does not answer this question.

On the other hand we worked out the model of associative memory, intended for implementation as an optical device ([13],[14]). Such a network is capable to hold and handle information that is encoded in the form of the frequency-phase modulation. In the network the signals propagate along interconnections in the form of quasi-monochromatic pulses at q different frequencies. There are arguments in favour of this idea. First of all, the frequency-phase modulation is more convenient for optical processing of signals. It allows us to back down an artificial adaptation of an optical network to amplitude modulated signals. Second, when signals with q different frequencies can propagate along one interconnection this is an analog of the channel multiplexing. In fact, this allows us to reduce the number of interconnections by a factor of q^2. Note that interconnections occupy nearly 98% of the area of neurochips.

In the center of our model the parametrical four-wave mixing process (FWM) is situated, that is well-known in nonlinear optics [15]. However, in order this model has good characteristics, an important condition must be added that should facilitate the propagation of useful signal, and, in the same time, suppress internal noise. This condition is *the principle of incommensurability of frequencies* proponed in [13],[14] in nonlinear optics terms (see Sec. 3).

The signal-noise analysis of our model made with the aid of the Chebyshev-Chernov statistical method [16],[17] showed that the storage capacity of the network was approximately q^2 times as much as the HM storage capacity. We called our network *the parametrical neural network* (PNN).

We worked out *the vector formalism* – universal description of PNN, not related directly to the optical model [18]-[20]. This formalism proved to be useful also for clear description of PGNN, although initially it was formulated in absolutely another terms. In this way one can easily establish relations between PGNN and PNN and also clarify the mechanisms, responsible for outstanding recognizing properties of both models. The reason is the local architecture of both networks, which suppresses system internal noise. In other q-valued

models there is no such suppression.

In this paper we give PGNN description, using the vector formalism. Then we define our PNN, using nonlinear optics terms and the vector formalism as well. Moreover, we consider some possible architectures for PNN.

Note. Our vector formalism is almost identical to the vector-neuron approach, which was suggested some years ago by [21]. We have found this paper after working out our own vector formalism. Dynamical rule in [21] was formulated not in the best way, however it seems, that the authors of [21] were the first to suggest the fruitful idea about representation of interconnections matrix as tensor product of vector-neurons.

II. Potts-glass neural network

We describe PGNN in terms of our vector formalism and in future compare it with PNN.

A. Vector formalism

PGNN consists of N neurons each of which can be in q different states. In order to describe the q different states of neurons we use the set of of q-dimensional vectors of a special type, so named *Potts vectors*. Namely, the lth state of a neuron is described by a column-vector $\mathbf{d}_l \in \mathrm{R}^{\mathrm{q}}$,

$$\mathbf{d}_l = \frac{1}{q}\begin{pmatrix} -1 \\ \vdots \\ q-1 \\ \vdots \\ -1 \end{pmatrix},\ l = 1,\ldots,q.$$

The state of the i-th neuron is described by a vector $\mathbf{x}_i = \mathbf{d}_{l_i}, 1 \le l_i \le q$. The state of the network as a whole X is determined by a set of N column-vectors $\mathbf{x}_i$: $X = (\mathbf{x}_1,\ldots,\mathbf{x}_N)$. The p stored patterns are

$$X^{(\mu)} = (\mathbf{x}_1^{(\mu)},\ldots,\mathbf{x}_N^{(\mu)}),\ \mathbf{x}_i^{(\mu)} = \mathbf{d}_{l_i^{(\mu)}},$$
$$1 \le l_i^{(\mu)} \le q,\ \mu = 1,2,\ldots,p.$$

Since neurons are vectors, the local field $\mathbf{h}_i$ affecting the ith neuron is a vector too,

$$\mathbf{h}_i = \frac{1}{N}\sum_{j=1}^{N}\mathbf{T}_{ij}\mathbf{x}_j.$$

The $(q\times q)$-matrices $\mathbf{T}_{ij}$ describe the interconnections between the ith and the jth neurons. By analogy with the Hopfield model these matrices are chosen in generalized Hebb form:

$$\mathbf{T}_{ij} = (1-\delta_{ij})\sum_{\mu=1}^{p}\mathbf{x}_i^{(\mu)}\mathbf{x}_j^{(\mu)+}, \quad i,j = 1,\ldots,N, \tag{1}$$

where $\mathbf{x}^+$ is q-dimensional row-vector and δ_{ij} is the Kronecker symbol. The matrix $\mathbf{T}_{ij}$ affects the vector $\mathbf{x}_j \in \mathrm{R}^{\mathrm{q}}$, converting it in a linear combination of column-vectors $\mathbf{d}_l$. After summation over all j we get the local field $\mathbf{h}_i$ as linear combination of vectors $\mathbf{d}_l$

$$\mathbf{h}_i = \sum_{l=1}^{q} A_l^{(i)}\mathbf{d}_l.$$

Let k be the index relating to the maximal coefficient: $A_k^{(i)} > A_l^{(i)}\ \forall\ l$. Then, by definition, the i-th neuron at the next time step, $t+1$, is oriented along a direction mostly close to the local field $\mathbf{h}_i$ at the time t:

$$\mathbf{x}_i(t+1) = \mathbf{d}_k. \tag{2}$$

The evolution of the system consists of consequent changes of orientations of vector-neurons according to the rule (2). We make the convention that if some of the coefficients $A_l^{(i)}$ are maximal simultaneously, and the neuron is in one of these *unimprovable* states, its state does not change. Then it is easy to show that during the evolution of the network its *energy* $H(t) = -1/2\sum_{i=1}^{N}(\mathbf{h}_i(t)\mathbf{x}_i(t))$ decreases. In the end the system reaches a local energy minimum. In this state all the neurons $\mathbf{x}_i$ are oriented in an unimprovable manner, and the evolution of the system come to its end. These states are the fixed points of the system. The necessary and sufficient conditions for a configuration X to be a fixed point is fulfillment of the set of inequalities:

$$(\mathbf{x}_i\mathbf{h}_i) \ge (\mathbf{d}_l\mathbf{h}_i), \quad \forall\ l = 1,\ldots,q;\ \forall\ i = 1,\ldots,N. \tag{3}$$

When $q = 2$, PGNN is the same as the standard Hopfield model.

B. Storage capacity of PGNN

Let we have the randomized patterns $\{X^{(\mu)}\}_1^p$. Suppose that the network starts from a distorted mth pattern

$$\tilde{X}^{(m)} = (\hat{b}_1\mathbf{x}_1^{(m)}, \hat{b}_2\mathbf{x}_2^{(m)},\ldots,\hat{b}_N\mathbf{x}_N^{(m)}).$$

The noise operator $\hat{b}_j$ with the probability b changes the state of the vector $\mathbf{x}_j^{(m)}$, and with the probability $1-b$ this vector remains unchanged. In other words, b is the probability of an error in a state of a neuron. The noise operators $\hat{b}_j$ are independent, too.

The network recognizes the reference pattern X^m correctly, if the output of the ith neuron defined by Eq.(2) is equal to $\vec{x}_i^{(m)}$. Otherwise, PGNN fails to recognize the pattern X^m. Let us estimate the probability of error in the recognition of mth pattern.

Simple calculations show, that probability of inequality validity $(\mathbf{x}_i^{(m)}\mathbf{h}_i) < (\mathbf{d}_l\mathbf{h}_i)$ at $\mathbf{d}_l \ne \mathbf{x}_i^{(m)}$ can be expressed as

$$\mathrm{Prob}\left\{\xi < \eta\right\} = \mathrm{Prob}\left\{\frac{1}{N}\sum_{j\ne i}^{N}\xi_j < \frac{1}{N}\sum_{j\ne i}^{N}\sum_{\mu\ne m}^{p}\eta_j^{(\mu)}\right\}, \tag{4}$$

where $\eta_j^{(\mu)} = (\mathbf{d}_l - \mathbf{x}_i^{(m)}, \mathbf{x}_i^{(\mu)})(\mathbf{x}_j^{(\mu)}\hat{b}_j\mathbf{x}_j^{(m)})$, $\xi_j = (\mathbf{x}_j^{(m)}\hat{b}_j\mathbf{x}_j^{(m)})$.

The quantity ξ is *the useful signal*. It is connected with influence of exactly the mth pattern onto the ith neuron. The partial random variables ξ_j are independent and identically distributed. The quantity η symbolizes *the inner noise*, connected with distorting influence of all other patterns. Partial noise

components $\eta_j^{(\mu)}$ are independent and identically distributed. It is easy to obtain the distributions for ξ_j and $\eta_j^{(\mu)}$:

$$\left\{\begin{array}{rcl} \xi_j & = & \left\{\begin{array}{rl} (q-1)/q, & 1-b \\ -1/q, & b \end{array}\right. , \\ \\ \eta_j^{(\mu)} & = & \left\{\begin{array}{rl} (q-1)/q, & 1/q^2 \\ 1/q, & (q-1)/q^2 \\ 0, & (q-2)/q \\ -1/q, & (q-1)/q^2 \\ -(q-1)/q, & 1/q^2 \end{array}\right. . \end{array}\right. \tag{5}$$

Let us pay attention on the fact, that at $q >> 1$ the noise component $\eta_j^{(\mu)}$ is localized mainly in zero:

$$\mathrm{Prob}\left\{\eta_j^{(\mu)} = 0\right\} = (q-2)/q \sim 1.$$

Total random variables ξ, η are asymptotic normal distributed with parameters

$$\begin{array}{ll} E(\xi) = \frac{q-1}{q} - b, & E(\eta) = 0, \\ D(\xi) \to 0; & D(\eta) = \frac{2(q-1)}{q^3} \cdot \alpha. \end{array} \tag{6}$$

where as usual *the loading parameter* $\alpha = \frac{p}{N}$. Now the probability of recognition error of coordinate $\mathbf{x}_i^{(m)}$ can be calculated by integration of the area under the "tail" of normally distributed η, where $\eta > E(\xi)$. Here we can explain, why the storage capacity of PGNN is much larger than HM.

The same considerations we can are valid for HM. It is done for example in [2]. Again we obtain a useful signal ξ and an internal noise η, and Eq. (4) for the probability of recognition failure. Again these random quantities will asymptotic normal as sums of independent, identically distributed partial random components ξ_j and $\eta_j^{(\mu)}$. The distributions of these last components can be obtained from Eq.(5) at $q = 2$ (because PGNN transforms into HM in this case). Mean values and dispersions for ξ and η can be obtained from (6) in the same way. As the result we have for HM:

$$\begin{array}{ll} \xi_j = \left\{\begin{array}{rl} 1/2, & 1-b \\ -1/2, & b \end{array}\right. , & \eta_j^{(\mu)} = \left\{\begin{array}{rl} 1/2, & 1/2 \\ -1/2, & 1/2 \end{array}\right. , \\ E(\xi) = \frac{1}{2} - b, & E(\eta) = 0, \\ D(\xi) \to 0; & D(\eta) = \frac{\alpha}{4}. \end{array} \tag{7}$$

Comparison of (7) with (5) and (6) demonstrates, that the dispersion of internal noise for PGNN is much smaller, than that for HM:

$$D_{PGNN}(\eta)/D_{HM}(\eta) = \frac{8(q-1)}{q^3} << 1, \text{ when } q >> 1.$$

Already at $q \sim 10$ the internal noise dispersion for PGNN is an order of magnitude smaller, than that for HM. Moreover, at $q \sim 10^2$ the fall of the dispersion is four orders of magnitude! This defines PGNN superiority over HM. We will give explanation of mechanism of internal noise compression in PGNN in the following Section.

Switching from one vector-coordinate situation to that with the whole pattern and using the standard approximation ([19],[20]) we obtain the expression for the probability of the error in the recognition of the pattern $X^{(m)}$,

$$\mathrm{Pr}_{err} \sim \sqrt{Np}\exp\left(-\frac{N}{2p}\frac{q(q-1)}{2}(1-\bar{b})^2\right), \quad \bar{b} = \frac{q}{q-1}b. \tag{8}$$

The expression sets the upper limit for the probability of recognition failure for PGNN. Then, the asymptotically possible value of the storage capacity of PGNN is

$$p_c = \frac{N}{2\ln N}\frac{q(q-1)}{2}(1-\bar{b})^2. \tag{9}$$

When $q = 2$, these expressions give the known estimates for HM. For $q > 2$ the storage capacity of PGNN is $q(q-1)/2$ times as large as the storage capacity of HM. In [3] the same factor was obtained by fitting the results of numerical calculations. We obtain the same result rigorously.

III. Parametrical neural network

Here we describe our associative memory model both in nonlinear optics and vector-formalism terms. We also will set out the obtained results for this model.

A. Nonlinear optic formulation

In the network the signals propagate along interconnections in the form of quasi-monochromatic pulses at q different frequencies

$$\{\omega_l\}_1^q \equiv \{\omega_1, \omega_2, ..., \omega_q\}. \tag{10}$$

The model is based on a parametrical neuron that is a cubic nonlinear element capable to transform and generate frequencies in the parametrical FWM-processes $\omega_i - \omega_j + \omega_k \to \omega_r$. Schematically this model of a neuron can be assumed as a device that is composed of a summator of input signals, a set of q ideal frequency filters $\{\omega_l\}^q$, a block comparing the amplitudes of the signals and q generators of quasi-monochromatic signals $\{\omega_l\}^q$.

Let $\{K^{(\mu)}\}_1^p$ be a set of patterns each of which is a set of quasi-monochromatic pulses with frequencies defined by Eq.(10) and amplitudes equal to ± 1:

$$\begin{array}{l} K^{(\mu)} = (\kappa_1^{(\mu)}, \ldots, \kappa_N^{(\mu)}),\ \kappa_i^{(\mu)} = \pm\exp(\imath\omega_{l_i^{(\mu)}}t), \\ \mu = 1, \ldots, p;\ i = 1, \ldots, N;\ 1 \le l_i^{(\mu)} \le q. \end{array} \tag{11}$$

The memory of the network is localized in interconnections $T_{ij},\ i,j = 1, \ldots, N$, which accumulate the information about the states of ith and jth neurons in all the p patterns. We suppose that the interconnections are dynamic ones and that they are organized according to the Hebb rule:

$$T_{ij} = (1-\delta_{ij})\sum_{\mu=1}^{p}\kappa_i^{(\mu)}\kappa_j^{(\mu)*}, \quad i,j = 1, \ldots, N. \tag{12}$$

The network operates as follows. A quasi-monochromatic pulse with a frequency ω_{l_j} that is propagating along the (ij)-th interconnection from the jth neuron to the ith one, takes part in FWM-processes with the pulses stored in the interconnection, $\omega_{l_i^{(\mu)}} - \omega_{l_j^{(\mu)}} + \omega_{l_j} \to \{\omega_l\}_1^q$. The amplitudes

±1 have to be multiplied. Summing up the results of these partial transformations over all patterns, $\mu = 1, \ldots, p$, we obtain a packet of quasi-monochromatic pulses, where all the frequencies from the set (10) are present. This packet is the result of transformation of the pulse ω_{l_j} by the interconnection T_{ij}, and it comes to the ith neuron. All such packets are summarized in this neuron. The summarized signal propagates through q parallel ideal frequency filters. The output signals from the filters are compared with respect to their amplitudes. The signal with the maximal amplitude activates the i-th neuron ('winner-take-all'). As a result it generates an output signal whose frequency and phase are the same as the frequency and the phase of the activating signal.

Generally, when three pulses interact, under a FWM-process always the fourth pulse appears. The frequency of this pulse is defined by the conservation laws only. However, in order that the abovementioned model works as a memory, an important condition must be add, which has to facilitate the propagation of the useful signal, and, in the same time, to suppress external noise. This condition is *the principle of incommensurability of frequencies* proponed in [13],[14]: *no combinations $\omega_l - \omega_{l'} + \omega_{l''}$ can belong to the set (10), when all the frequencies are different.*

Now we finished to describe the principle of the network operating. This network will be called *the parametrical neural network* (PNN). Here an important remark has to be done.

Generally speaking, there are different parametrical FWM-processes complying with the principle of incommensurability of frequencies. However, better results can be obtained for the parametrical FWM-process

$$\omega_l - \omega_{l'} + \omega_{l''} = \begin{cases} \omega_l, & \text{when } l' = l''; \\ \to 0, & \text{in other cases.} \end{cases} \quad (13)$$

This architecture will be called PNN-2 (another architecture, PNN-1, was examined in [13],[14]). Here we investigate the abilities of PNN-2. The structure of the rest of the paper is as follows. In next subsection we introduce a vector formalism allowing us to formulate the problem in the general form. Then, the results for PNN-2 will be presented. Then we mention shortly about other neuro-architectures, based on PNN-2. Some remarks are given in Conclusions.

B. Vector formalism for PNN-2

In order to describe the q different states (10) of neurons we use the set of basis vectors $\mathbf{e}_l$ in the space $\mathrm{R}^{\mathrm{q}},\ \mathrm{q} \geq 1$,

$$\mathbf{e}_l = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix},\ l = 1, \ldots, q.$$

The state of the ith neuron is described by a vector $\mathbf{x}_i$,

$$\mathbf{x}_i = x_i \mathbf{e}_{l_i},\ x_i = \pm1,\ \mathbf{e}_{l_i} \in \mathrm{R}^{\mathrm{q}}, \begin{cases} 1 \leq l_i \leq q; \\ i = 1, \ldots, N. \end{cases} \quad (14)$$

The factor x_i denotes the signal phase. The state of the network as a whole X is determined by a set of N q-dimensional vectors $\mathbf{x}_i$: $X = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$. The p stored patterns are

$$X^{(\mu)} = (\mathbf{x}_1^{(\mu)}, \mathbf{x}_2^{(\mu)}, \ldots, \mathbf{x}_N^{(\mu)}),\ \mathbf{x}_i^{(\mu)} = x_i^{(\mu)} \mathbf{e}_{l_i^{(\mu)}},$$
$$x_i^{(\mu)} = \pm1,\ 1 \leq l_i^{(\mu)} \leq q,\ \mu = 1, \ldots, p,$$

and the local field is

$$\mathbf{h}_i = \frac{1}{N} \sum_{j=1}^{N} \mathbf{T}_{ij} \mathbf{x}_j. \quad (15)$$

The $(q \times q)$-matrix $\mathbf{T}_{ij}$ describes the interconnection between the ith and the jth neurons. This matrix affects the vector $\mathbf{x}_j \in \mathrm{R}^{\mathrm{q}}$, converting it in a linear combination of basis vectors $\mathbf{e}_l$. This combination is an analog of the packet of quasi-monochromatic pulses that come from the jth neuron to the ith one after transformation in the interconnection. To satisfy the conditions (12) and (13), we need to take the matrices $\mathbf{T}_{ij}$ as

$$\mathbf{T}_{ij} = (1 - \delta_{ij}) \sum_{\mu=1}^{p} \mathbf{x}_i^{(\mu)} \mathbf{x}_j^{(\mu)+}, \quad i, j = 1, \ldots, N. \quad (16)$$

Note, that the structure of this expression is similar to that of (1).

The dynamic rule is left as earlier: the ith neuron at the time $t+1$ is oriented along a direction mostly close to the local field $\mathbf{h}_i(t)$. However the expressions will differ from (2). Indeed, with the aid of (16) we write $\mathbf{h}_i$ in the form more convenient for analysis:

$$\mathbf{h}_i(t) = \sum_{l=1}^{q} A_l^{(i)} \mathbf{e}_l,\ A_l^{(i)} \sim \sum_{j(\neq i)}^{N} \sum_{\mu=1}^{p} (\mathbf{e}_l \mathbf{x}_i^{(\mu)})(\mathbf{x}_j^{(\mu)} \mathbf{x}_j(t)). \quad (17)$$

Let k be the index relating to the amplitude that is maximal *in modulus* in the series (17): $\mid A_k^{(i)} \mid > \mid A_l^{(i)} \mid \quad \forall\ l$. Then according to our definition,

$$\mathbf{x}_i(t+1) = \mathrm{sgn}(A_k^{(i)}) \mathbf{e}_k. \quad (18)$$

The evolution of the system consists of consequent changes of orientations of vector-neurons according to the rule (18). The necessary and sufficient conditions for a configuration X to be a fixed point is fulfillment of the set of inequalities:

$$(\mathbf{x}_i \mathbf{h}_i) \geq \mid (\mathbf{e}_l \mathbf{h}_i) \mid, \quad \forall\ l = 1, \ldots, q;\ \forall\ i = 1, \ldots, N,$$

(compare with Eq.(3)).

C. Storage capacity of PNN-2

All these considerations are identical to those for PGNN. Differences appear only because of neurons are defined now not only by vectors, but also by scalars ±1. The distorted mth pattern has the form

$$\tilde{X}^{(m)} = (a_1 \hat{b}_1 \mathbf{x}_1^{(m)}, a_2 \hat{b}_2 \mathbf{x}_2^{(m)}, \ldots, a_N \hat{b}_N \mathbf{x}_N^{(m)}).$$

Here $\{a_i\}_1^N$ and $\{\hat{b}_i\}_1^N$ define a *phase noise* and a *frequency noise* respectively: a_i is a random value that is equal to -1

or $+1$ with the probabilities a and $1-a$ respectively; b is the probability that the operator $\hat{b}_i$ changes the state of the vector $\mathbf{x}_i^{(m)} = x_i^{(m)}\mathbf{e}_{l_i^{(m)}}$, and $1-b$ is the probability that this vector remains unchanged.

The amplitudes $A_l^{(i)}$ (17) have the form

$$A_l^{(i)} \sim \begin{cases} x_i^{(m)} \sum_{j\neq i}^N \xi_j \quad + \quad \sum_{j\neq i}^N \sum_{\mu\neq m}^p \eta_j^{(\mu)}, \quad l = l_i^{(m)}; \\ \qquad\qquad\qquad \sum_{j\neq i}^N \sum_{\mu\neq m}^p \eta_j^{(\mu)}, \quad l \neq l_i^{(m)}, \end{cases}$$

where $\eta_j^{(\mu)} = a_j(\mathbf{e}_l\mathbf{x}_i^{(\mu)})(\mathbf{x}_j^{(\mu)}\hat{b}_j\mathbf{x}_j^{(m)})$, $\xi_j = a_j(\mathbf{x}_j^{(m)}\hat{b}_j\mathbf{x}_j^{(m)})$, $j(\neq i) = 1,\ldots,N$, $\mu(\neq m) = 1,\ldots,p$. When the patterns $\{X^{(\mu)}\}_1^p$ are uncorrelated, the quantities ξ_j and $\eta_j^{(\mu)}$ are independent random variables described by the probability distributions

$$\xi_j = \begin{cases} +1, & (1-b)(1-a) \\ 0, & b \\ -1 & (1-b)a \end{cases}, \quad \eta_j^{(\mu)} = \begin{cases} +1, & 1/2q^2 \\ 0, & 1-1/q^2 \\ -1 & 1/2q^2 \end{cases},$$

(compare with Eq.(5)). As in the case of PGNN, when $q >> 1$ the noise component $\eta_j^{(\mu)}$ is localized mainly in zero:

$$\text{Prob}\left\{\eta_j^{(\mu)} = 0\right\} = 1 - 1/q^2 \sim 1.$$

Eq.(6) now will transform into:

$$\begin{array}{ll} E(\xi) = (1-2a)(1-b), & E(\eta) = 0, \\ D(\xi) \to 0; & D(\eta) = \frac{1}{q^2}\cdot\alpha. \end{array}$$

When $q >> 1$ the dispersion of internal noise for PNN-2 is even smaller, than for PGNN:

$$D_{PNN}(\eta)/D_{PGNN}(\eta) = 1/2, \text{ when } q >> 1.$$

In the long run this determines the superiority of PNN-2 over PGNN in memory capacity and noise immunity. It is convenient here to mention mechanisms, suppressing internal noises. They are identical in both models, but we will demonstrate them on the PNN example.

When signal propagates it interacts with frequencies, stored in interconnection $\omega_{l_i^{(\mu)}} - \omega_{l_j^{(\mu)}} + \omega_{l_j} \to \{\omega_l\}_1^q$. In addition the principal of frequencies incommensurability (13) should be fulfilled. It can be formulated in vector terms as:

$$\mathbf{x}_i^{(\mu)}\mathbf{x}_j^{(\mu)+}\mathbf{x}_j = \begin{cases} \mathbf{x}_i^{(\mu)}, & \text{when } l_j^{(\mu)} = l_j; \\ 0, & \text{in other cases.} \end{cases}$$

One can see from the last equation, that the largest part of propagated signals will be suppressed. It happens because the interconnection chooses the only one combinations of indices $l_j^{(\mu)}$ and l_j from all possible ones, where indices coincide (other combinations give zero). In other words, the interconnection filters signals. It is the main reason of the largest part of internal noise η is localized in zero.

The similar filtration happens also in PGNN. The difference is that in PGNN the signal always propagates through the interconnection. But when indices $l_j^{(\mu)}$ and l_j coincide, the signal is attributed with large positive amplitude ~ 1. If indices do not coincide, the signal is attributed with small negative amplitude $\sim -1/q$. This signal filtration leads to suppression of internal noise in PGNN. In all another q-valued models of associative memory this filtration is absent.

At the end of consideration of PNN-2 we give the expressions for noise immunity and storage capacity similar to (8) and (9):

$$\text{Pr}_{err} \sim \sqrt{Np}\exp\left(-\frac{N(1-2a)^2}{2p}\cdot q^2(1-b)^2\right), \quad (19)$$

$$p_c = \frac{N(1-2a)^2}{2\ln N}\cdot q^2(1-b)^2. \quad (20)$$

When $q = 1$, Eqs.(19)-(20) transform into well-known results for the standard Hopfield model (in this case there is no frequency noise, $b = 0$). When q increases, the probability of the error (19) decreases exponentially, i.e. the noise immunity of PNN increases noticeably. In the same time the storage capacity of the network increases proportionally to q^2. In contrast to the Hopfield model the number of the patterns p can be much greater than the number of neurons.

For example, let us set a constant value $\text{Pr}_{err} = 0.01$. In the Hopfield model, with this probability of the error we can recognize any of $p = N/10$ patterns, each of which is less then 30% noisy. In the same time, PNN-2 with $q = 64$ allows us to recognize any of $p = 5N$ patterns with 90% noise, or any of $p = 50N$ patterns with 65% noise. Our computer simulations confirm these results.

The memory capacity in PNN-2 is twice as large as that in PGNN. Evidently, it is connected with the fact, that for the same q the number of different states of neurons in PNN-2 is twice as large as that in PGNN. In general, both models have very similar characteristics.

D. Other PNN-architectures

1) Phase-independent PNN-3: When the PNN is realized as a device, the problem arises, that one should control the phases of all signals. All phases should be matched. It is rather difficult problem. It seems, that the easiest way to overcome this difficulty is to make all phases identical. Formally, we should make all amplitudes ± 1 in (11) and (14) to 1. More precise analysis shows, that in this case partial noise components $\eta_j^{(\mu)}$ become not independent. The noise dispersion drastically increases. The way out is to use specially chosen vector thresholds in the local field definition [22]:

$$\mathbf{h}_i = \frac{1}{N}\left\{\sum_{j=1}^N \mathbf{T}_{ij}\mathbf{x}_j - \frac{1}{q}\sum_{\mu=1}^p \mathbf{x}_i^{(\mu)}\right\}, \quad (21)$$

where matrices $\mathbf{T}_{ij}$ are determined by Eq.(16). Then the partial noise components $\eta_j^{(\mu)}$ become uncorrelated. And it is possible to apply the probability-theoretic approach for estimation of signal/noise ratio.

Means and dispersions of total random variables ξ and η are the same as in expressions (6). But whole phase-independent PNN (we called it as PNN-3) is equivalent to PGNN. If to compare with PNN-3-model PGNN is too complicated. It is

related with using the Potts vectors $\mathbf{d}_l$ instead of basis vectors $\mathbf{e}_l$. Being realized as a computer algorithm PNN-3 works q times quicker than PGNN.

2) Decorrelating PNN: We suggested the method of sufficient enlarging of binary associative memory with the help of PNN-architecture for the case of correlation between patterns ([23],[24]). As it is known the memory capacity of Hopfield model falls down drastically if there are correlations, so the only way out is so named *sparse coding* [25]-[29]. Our method is an alternative to this approach.

At the heart of our approach is one-to-one mapping of binary patterns into internal representation, using vector-neurons of large dimension, $q >> 1$. Then PNN is being constructed on the basis of obtained vector-neuron patterns. The representation has the following properties: *i)* correlations between vector-neuron patterns become negligible; *ii)* dimension q of vector-neurons increases *exponentially* as a function of mapping parameter. The larger a dimension q the better recognition properties of PNN. The result of exponential increase of q leads to the exponential increase of binary memory capacity.

The mapping of binary patterns into vector-neuron ones is based on the very clear idea. This idea resembles the method, which was used previously in sparse coding ([30]), where due to a redundant coding it was possible to increase the storage capacity comparing with the Hopfield model. In the same time the noise immunity of the system was very low. In our case the redundancy of coding is absent, the storage capacity increases drastically, and the noise immunity is much greater. In future we plan to compare PNN with sparse coding in details.

IV. Conclusions

From the early 90th the intensity of q-valued neural networks researches sharply decreased. Presumably it can be explained by absence of progress in development of effective models of associative memory. Computer algorithm of PNN-architecture demonstrates, that we approach to those magnitudes of storage capacity and noise immunity which could be of interest for practical applications. Use of PNN-architectures seems to us very promising.

Acknowledgment

The authors would like to thank Dr. Tatyana Lisovsky for her help.

This work was supported by the program "Intellectual Computer Systems" (the project 2.45) and in part by Russian Basic Research Foundation (grants 02-01-00457 and 03-01-00355)

References

[1] D. Amit, H. Gutfreund, H. Sompolinsky, "Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks," *Phys. Rev. Lett.*, vol. 55, pp. 1530-1533, 1985; "Information storage in neural networks with low levels of activity," *Phys. Rev. A*, vol. 35, pp. 2293-2303, 1987.
[2] J. Hertz, A. Krogh, R. Palmer, *Introduction to the Theory of Neural Computation*, NY: Addison-Wesley, 1991.
[3] I. Kanter, "Potts-glass models of neural networks," *Physical Review A*, vol. 37, pp. 2739-2742, 1988.
[4] J. Noest, "Discrete-state phasor neural networks," *Physal Review A*, vol. 38, pp. 2196-2199, 1988.
[5] J. Cook, "The mean-field theory of a Q-state neural network model," *Journal of Physics A*, vol. 22, pp. 2000-2012, 1989.
[6] H. Rieger, "Storing an extensive number of grey-toned patterns in a neural network using multistate neurons," *Journal of Physics A*, vol. 23, pp. L1273-L1279, 1990.
[7] D. Bolle, P. Dupont, J. van Mourik, "Stability properties of Potts neural networks with biased patterns and low loading," *Journal of Physics A*, vol. 24, pp. 1065-1081, 1991.
[8] H. Vogt, A. Zippelius, "Invariant recognition in Potts glass neural networks," *Journal of Physics A*, vol. 25, pp. 2209-2226, 1992.
[9] D. Bolle, P. Dupont, J. Huyghebaert, "Thermodynamics properties of the q-state Potts-glass neural network," *Physical Rewiew A*, vol. 45, pp. 4194-4197, 1992.
[10] D. Bolle, J. van Mourik, "Capacity of diluted multi-state neural networks," *Journal of Physics A*, vol. 27, pp. 1151-1162, 1994.
[11] F.Y Wu, "The Potts model," *Review of Modern Physics*, vol. 54, pp. 235-268, 1982.
[12] R. Baxter, *Exactly Solved Models in Statistical Mechanics*, London: Academic Press, 1982.
[13] B.V. Kryzhanovsky, A.L. Mikaelian, "On recognition ability of neuronet based on neurons with parametrical frequencies convertion," *Doklady Mathematics*, vol. 65(2), pp. 286-288, 2002.
[14] A. Fonarev, B.V. Kryzhanovsky et al., "Parametric dynamic neural network recognition power," *Optical Memory & Neural Networks*, vol. 10(4), pp. 31-48, 2001.
[15] N. Bloembergen, *Nonlinear optics*, NY: Benjamin, 1965.
[16] N. Chernov, "A mesure of asymptotic efficiency for tests of hypothesis based on the sum of observations," *Ann. Math. Statistics*, vol. 23, pp. 493-507, 1952.
[17] B.V. Kryzhanovsky, A.L. Mikaelian, V.N. Koshelev, et al., "On recognition error bound for associative Hopfield memory," *Optical Memory & Neural Networks*, vol. 9(4), pp. 267-276, 2000.
[18] B.V. Kryzhanovsky, L.B. Litinskii, A. Fonarev, "Optical neural network based on the parametrical four-wave mixing process," In: Wang, L., Rajapakse, J.C, Fukushima, K., Lee, S.-Y., and Yao, X. (eds.): *Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02)*, Orchid Country Club, Singapore, vol. 4, pp. 1704-1707, 2002.
[19] B.V. Kryzhanovsky, L.B. Litinskii, "Vector models of associative memory," *Automaton and remote control*, vol. 64(11), pp. 1782-1793, 2003.
[20] A.L. Mikaelian, B.V. Kryzhanovsky, L.B. Litinskii, "Parametrical Neural Network," *Optical Memory & Neural Networks*, vol. 12(3), pp. 227-236, 2003.
[21] Y. Nakamura, K. Torii, T. Munaka, "Neural-network model composed of multidimensional spin neurona," *Phys. Rev. B*, vol. 51, pp. 1538-1546, 1995.
[22] D.I. Alieva, B.V. Kryzhanovsky, "Phaseless parametrical neural net," *International Conference on Artificial Intelligent Systems IEEE AIS'2003*, vol. 1, pp. 511-517, Moscow, Fizmatlit, 2003 (in russian).
[23] B.V. Kryzhanovsky, A.L. Mikaelian, "An associative memory capable of recognizing strongly correlated patterns," *Doklady Mathematics*, vol. 67(3), pp. 455-459, 2003.
[24] B. Kryzhanovsky, L. Litinskii, and A. Fonarev, "An effective associative memory for pattern recognition," In: *Advances in Intelligent Data Analysis V. 5th International Symposium on Intelligent Data Analysis IDA 2003*, pp. 179-186, Berlin: Springer, 2003.
[25] G. Palm, "Memory capacity of local rules for synaptic modification," *Concepts in Neuroscience*, vol. 2, pp. 97-128, 1991.
[26] G. Palm, "On the information storage capacity of local learning rules," *Neural Computation*, vol. 4, pp. 703-711, 1992.
[27] G. Palm, F.T. Sommer, "Information capacity in recurrent McCulloch-Pitts networks with sparsely coded memory states," *Network*, vol. 3, pp. 1-10, 1992.
[28] C.J. Perez-Vicente, D.J. Amit, "Optimized network for sparsely coded patterns," *Journal of Physics A*, vol. 22, pp. 559-569, 1989.
[29] A.A. Frolov, I.P. Murav'ev, "Informational characteristics of neural networks capable of associative learning based on Hebbiean plasticity," *Network*, vol. 4, pp. 495-536, 1995.
[30] G. Palm, 2003, private comunication.